

A room-temperature noise-free quantum memory for broadband light

K. T. Kaczmarek^{1,*}, P. M. Ledingham¹, B. Brecht¹, S. E. Thomas^{1,2}, G. S. Thekkadath³, O. Lazo-Arjona¹, J. H. D. Munns^{1,2}, E. Poem^{1,4}, A. Feizpour¹, D. J. Saunders¹, J. Nunn¹, I. A. Walmsley^{1*}

¹*Clarendon Laboratory, University of Oxford, Parks Road, Oxford, OX1 3PU, UK.*

²*QOLS, Blackett Laboratory, Imperial College London, London SW7 2BW, UK.*

³*University of Ottawa, 25 Templeton St, Ottawa, K1N 6N5, Canada.*

⁴*Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 7610001, Israel.*

(Dated: April 4, 2017)

Light can carry quantum information in ambient conditions. It also has a large information capacity by virtue of its high bandwidth, making it ideal for building quantum networks [1]. However, due to the probabilistic nature of linear-optical entangling operations, it remains an outstanding challenge to grow such networks [2]. Historically, the goal of swapping entanglement over extended distances motivated the development of “quantum repeaters”, based on quantum memories that can trap and release photons on demand to synchronise entangling operations over lossy networks [3]. In contrast to long-distance communications networks, local networks can operate fast, with low latency. Entangling operations could thus be synchronised using memories with shorter storage times. For quantum applications, the key figures of merit are low noise and the time-bandwidth product [4], i.e. the number of clock-cycles over which a memory allows synchronisation. While quantum-limited performance [5–9], large time-bandwidth products [10], and operation in ambient conditions [11] have been shown separately, to date no system satisfies all these desiderata simultaneously. Here we introduce and demonstrate a new memory protocol – the off-resonant cascaded absorption (ORCA) memory – that is optimised for low-latency applications, such as local synchronisation of multiple photons over short distances. Based on off-resonant stimulated two-photon absorption into a doubly excited electronic state in warm atomic vapour, ORCA is the first broadband and intrinsically noise-free room-temperature quantum memory. We use the ORCA scheme to store, and retrieve on-demand, GHz-bandwidth heralded single photons with 15% efficiency, measuring a heralded autocorrelation $g_h^{(2)} = 0.028(9)$, the most statistically significant anti-bunching recorded from any quantum memory. These results open a route to arbitrarily scalable local photonic quantum networks.

The basic operating principle of a quantum optical memory is to prepare an empty storage state $|s\rangle$ and then map on-demand a quantum state of light to and from an atomic coherence between the ground state $|g\rangle$ and $|s\rangle$. With an additional intermediate atomic level $|e\rangle$ and an auxiliary control field this can be achieved – the input “signal” field drives the “populated transition” ($|g\rangle \leftrightarrow |e\rangle$), while the control field drives the “empty transition” ($|e\rangle \leftrightarrow |s\rangle$), as shown in Fig. 1. By using atomic ensembles, one increases the probability of a photon interacting with an atom, leading to strong light-matter coupling [12]. Traditionally the energy levels are arranged in a Λ configuration: two ground states coupled via an excited state, since the long ground state coherence time gives prospects for long-distance quantum communication applications [13]. Narrowband (\sim MHz), resonant Λ -type quantum optical memories based on electromagnetically-induced transparency (EIT) [5, 14], gradient echo memory (GEM) [15] and the full atomic frequency comb (AFC) [16] protocol have performed well, with high efficiency [17] and preservation of photon statistics [18] being demonstrated. However, their narrowband nature combined with the additional need for preparation of the storage state (by means of optical pumping or spectral holeburning) limits their maximum clock-rate. Focus-

ing on the requirements of local quantum networks, being able to operate at a high clock-rate for elementary building block synchronisation motivates the need for a broadband memory for very short pulses of light. For instance, interfacing heralded single-photon sources with memories having just a 100 ns storage time and a storage bandwidth of 10 GHz, i.e. a time-bandwidth product of 1000, it should be possible to produce heralded 10-photon states at kHz rates compared to once a week without memories [4].

The Λ configuration does allow the freedom of having the signal and control fields far-off single-photon resonance while remaining in two-photon resonance, as in the Raman memory scheme [19, 20]. In this protocol, the control creates a broadband “virtual” resonance that the signal couples to. Because the signal is far detuned from the “real” atomic resonance, there is very little linear absorption. Interestingly, this means the system can be reduced to a beamsplitter-like interaction between different travelling optical modes and a stationary matter mode [21–23]. The caveat is that, in the absence of atomic selection rules, the control can off-resonantly couple to the populated transition, inducing spontaneous Raman scattering and polluting the storage state with spurious atomic excitations [24]. The latter leads to four-wave mixing (FWM) noise in the same optical mode as the desired quantum signal when the memory is read out, deteriorating its photon statistics [10, 24]. Noise con-

* k.kaczmarek1@physics.ox.ac.uk

tamination of the signal, due to control field induced fluorescence or especially FWM in off-resonant broadband memories, remains one of the greatest challenges prohibiting network ready quantum memories.

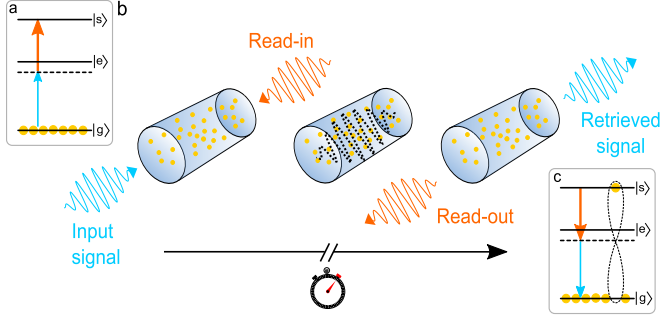


FIG. 1. Schematic representation of the ORCA protocol. **a**, Level diagram with the storage process shown. A weak signal field (blue) and strong control (orange) are on two-photon resonance with a doubly-excited storage state $|s\rangle$, while remaining detuned from the intermediate state $|e\rangle$. This configuration maps the signal field to an atomic coherence between the ground $|g\rangle$ and storage $|s\rangle$ states, as shown by the dotted line in **c**. **b**, The signal and read-in control pulses are overlapped counter-propagating in an atomic vapour, leading to absorption of the signal and formation of an atomic coherence. When a second, read-out control pulse is sent into the memory, it leads to re-emission of the signal in the forward direction. **c**, Level diagram with the retrieval process shown. Re-applying the control pulse leads to re-mapping of the atomic coherence back into an optical field at the signal frequency.

Here we introduce a new quantum memory scheme based on off-resonant cascaded absorption (ORCA) in a three-level ladder configuration that is optimised for warm atomic vapours (see Fig. 1). The operational principle is similar to the Raman memory: two broadband optical fields, signal and control, satisfy a two-photon resonance condition with a storage state, while individually being far-off resonance from their respective single-photon transitions. Similarly to Raman storage, unit memory efficiency can be achieved with appropriate shaping of the control pulse [25, 26]. However, unlike in Λ -type protocols, the memory bandwidth is not limited by the ground state splitting, enabling storage of >10 GHz bandwidth pulses. Furthermore, a counter-propagating beam geometry reduces motional-induced dephasing of the stored coherence, enabling ~ 100 ns storage times in warm vapours.

In the new scheme, the storage state is a doubly-excited electronic state, which has no thermal excitations even at high temperatures. Therefore the protocol requires no preparation of the atomic ensemble prior to storage, and there is no contamination of the retrieved fields due to imperfect optical pumping. This points to the major benefit of ORCA storage in that it is *noise-free*. The signal and control wavelengths can be chosen such that the control photons are significantly de-

tuned from the populated transition (THz detunings are readily available in the rich atomic structure of alkalis). This effectively eliminates any control field induced fluorescence noise. More importantly though, due to the ladder configuration, there is no scattering process that could populate the storage state, and so four-wave mixing noise is eliminated. Finally, efficient suppression of control field leakage on the output detection is readily achievable using off-the-shelf low-loss interference filters, and bolstered by the counter-propagating field configuration, in principle enabling very high *end-to-end* device efficiency. The ORCA scheme is therefore a broadband and noise-free quantum optical memory protocol operating within a technically simple, room-temperature platform, which to-date no other protocol offers.

As a proof-of-principle demonstration, we implement the protocol with near-infrared light in warm caesium vapour (see Methods section for more details). We interface the ORCA memory with a 1 GHz bandwidth heralded single-photon source based on parametric down-conversion (PDC) [24], as shown in Fig. 2. In type-II PDC, a “pump” field is converted into orthogonally polarised pairs of photons by means of spontaneous scattering. In the low-gain regime, when the pair production rate is low, the detection of a photon in one of the modes (“idler”) heralds the presence of another (“signal”). Our heralding efficiency, i.e. the probability of a single photon being present prior to storage given a herald detection event, is $\eta_{\text{herald}} \approx 5\%$. We store these single signal photons in ORCA for 3.5 ns, which is three orders of magnitude longer than the current state of the art of storing PDC photons in room temperature quantum memories [10].

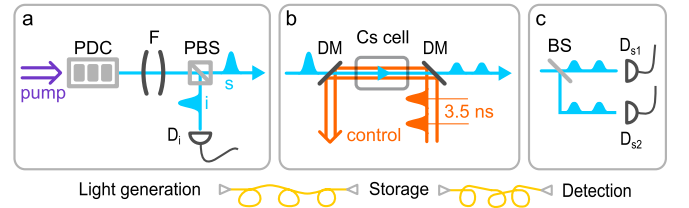


FIG. 2. Experimental setup. **a**, Photon pairs at 852 nm are generated in parametric down-conversion (PDC) and consecutively filtered (F) to ~ 1 GHz bandwidth. The signal (s) and idler (i) photons are separated on a polarizing beam splitter (PBS). While the idler photons are detected with a single-photon detector (D_i), the signal photons are sent to the memory. **b**, Dichroic mirrors (DM) are used to combine the signal field with a bright, counter-propagating control field at 917 nm (read-in and read-out) inside a caesium vapour cell. After the memory, the transmitted and retrieved light is sent to the detection stage. **c**, At the detection stage, the signal photons are split on a balanced beam splitter (BS) and detected by two single-photon detectors (D_{s1} and D_{s2}) in a Hanbury-Brown-Twiss configuration.

Fig. 3a shows a section of the photon arrival time trace with the control field off (“SIG”) and on (“MEM”). From this we extract a memory efficiency $\eta = 14.6 \pm 1.9\%$. The

retrieved pulse is longer than the input due to a small mismatch between signal and control bandwidths. The “CTRL” setting on Fig. 3a corresponds to the control field being on, but with no input signal field. Any photons detected in the “CTRL” configuration would correspond to noise generated by the memory itself. A convenient metric to measure noise performance of optical quantum memories is $\mu_1 = \langle n^{\text{noise}} \rangle / \eta$, i.e. the ratio of the average number of noise photons per control pulse $\langle n^{\text{noise}} \rangle$ and the memory efficiency η [27]. In our experiment, we find $\mu_1 \leq (39.4 \pm 0.2) \times 10^{-6}$, significantly lower than our η_{herald} , as required for single-photon storage. To our knowledge this is the lowest μ_1 ever reported. In fact, this is only an upper estimate for the actual μ_1 , since the measured noise is upper-bounded by the technical noise (dark counts) of our detectors.

To verify the quantum performance of the memory we measure the photon statistics of the retrieved signal and compare it with the single photons sent into the memory. Without the need to prepare the storage state, we are able to operate the experiment at the full 80 MHz repetition rate of our laser system. Fig. 3b shows the detected coincidence clicks between the detectors D_i & $D_{s1/2}$ at different times with the control off (“SIG”) and on (“MEM”). First, we evaluate the cross-correlation function $g^{(1,1)}$ of signal and idler photons (see Methods section). We measure $g^{(1,1)} = 131.3 \pm 0.2$ for the input signal field, and after storage and retrieval we obtain $g^{(1,1)} = 120.0 \pm 0.1$, clearly exceeding the classical bound of $g_{\text{classical}}^{(1,1)} \leq 2$. We attribute the reduction of the $g^{(1,1)}$ in the read-out bin to dark count contamination. Nevertheless, this explicitly demonstrates that ORCA preserves the non-classical correlations between the signal and idler.

As an even stronger benchmark of quantum performance, we evaluate the second-order heralded autocorrelation function $g_h^{(2)}$ (see Methods section). A value of $g_h^{(2)} < 1$ is a direct measure of anti-bunched photon statistics, with $g_h^{(2)} = 0$ corresponding to a perfect single photon. Figures 3c and 3d show the clearly anti-bunched nature of the input/retrieved fields. We measure the $g_h^{(2)}$ of our input field to be 0.020 ± 0.005 , well below the classical boundary of 1, as expected from low-gain PDC. We then measure the heralded auto-correlation function of the retrieved light, observing $g_h^{(2)} = 0.028 \pm 0.009$, 108 standard deviations below the classical limit and to our knowledge the lowest ever measured from a room-temperature quantum memory. The excellent agreement between input and output $g_h^{(2)}$ further confirms the memory generates no noise.

We have introduced the first noise-free broadband optical quantum memory protocol at room temperature, based on off-resonant cascaded absorption (ORCA), and used it to successfully store and retrieve GHz-bandwidth heralded single photons preserving their quantum statistics. We achieved the most statistically significant anti-bunching of photons yet measured from any quantum

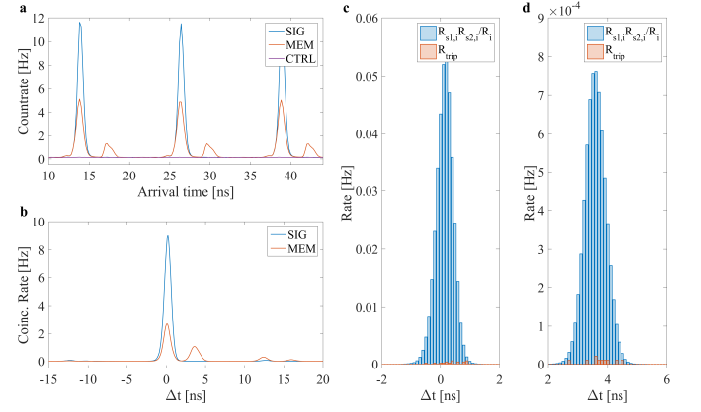


FIG. 3. Storage of single photons using ORCA. **a**, Histogram showing timing of total signal detector clicks with respect to external 1 MHz trigger. “SIG” is the signal field with the control field off, “MEM” is the signal field with the control on for a storage time of 3.5 ns. “CTRL” has the control field on, but with no input signal, demonstrating the noise-free nature of ORCA. **b**, Histogram of the time difference Δt between idler and total signal detector clicks, with the control off (“SIG”) and on (“MEM”). The peaks at 12.5 and 16 ns in the “MEM” trace come from the next series of read-in/-out pulses of the 80 MHz pulse train hitting the memory and reading out residual coherence from the atoms, similarly to [21]. **c**, Input signal field: orange - histogram of the time difference Δt between idler detector clicks, and signal detector coincidences, i.e. triple coincidence histogram R_{trip} ; blue - product of the two-fold coincidences between the idler and two signal detectors, $R_{s1,i}$ and $R_{s2,i}$, normalized by the idler counts R_i , i.e. predicted triple coincidence histogram for coherent light of the same average photon rate as the PDC. **d**, Same as **c**, but for the retrieved field. The ratio of the orange and the blue is the observed heralded autocorrelation $g_h^{(2)}$, with both **c** and **d** showing significant anti-bunching.

memory. The memory lifetime in the current implementation in Cs vapour was 5.4 ns, limited by motional-induced dephasing (see Supplementary Material), giving a time-bandwidth product ~ 5 . By using a different atomic species with a smaller signal/control wavevector mismatch (e.g. rubidium), the lifetime can be significantly extended, allowing one to attain time-bandwidth products > 1000 . Furthermore, high efficiencies can be achieved by appropriate shaping of the control pulse [25, 28]. These improvements will make it possible to reach the performance necessary for high clock-rate synchronisation of probabilistic operations in local quantum networks [4]. This will enable the creation of large photonic quantum states, opening the way to a new regime of quantum simulation, computation and sensing.

METHODS

Experiment

We use the Cs D2 line at 852 nm for our signal field, with $6S_{1/2}(F = 4)$ as the ground state $|g\rangle$ and the $6P_{3/2}(F = 3, 4, 5)$ manifold as the ORCA intermediate state $|e\rangle$. A strong 917 nm control field couples this signal to the storage state $|s\rangle$, i.e. the $6D_{5/2}(F = 2, 3, 4, 5, 6)$ manifold. It has been shown that such a complex atomic structure can be reduced to an effective three-level system for the case of broadband excitation [29]. We detune both fields by 6 GHz from the intermediate state towards the ground state, enabling good coupling with negligible linear absorption.

The generation of heralded single photons is achieved using type-II parametric down-conversion in a periodically poled potassium titanyl phosphate waveguide. The source generates THz-bandwidth pairs of signal and idler photons, both of which are consequently filtered down to ~ 1 GHz bandwidth centred at the signal frequency using a series of Fabry-Perot etalons and grating filters [24]. We send the PDC idler to a single-photon silicon avalanche photodiode (APD) and the signal towards the memory. Our heralding efficiency before the memory is $\eta_{\text{herald}} \approx 5\%$. The source is pumped at a rate of 80 MHz from a frequency doubled actively mode-locked titanium sapphire laser, with a $\sim 0.8\%$ chance of producing a photon pair of the correct bandwidth per pump pulse.

The control field is derived from a second mode-locked titanium sapphire laser, locked-to-clock with the PDC pump. The laser bandwidth is ~ 1 GHz, corresponding to a pulse duration of ~ 440 ps. In order to investigate storage times < 12.5 ns, i.e. smaller than the time between consecutive pulses from the laser, we use an unbalanced Mach-Zehnder interferometer with a variable delay in one arm to split the control pulse train into two: read-in and read-out, and delay them with respect to each other. The read-in and read-out control pulse energies are 0.21(1) and 0.97(1) nJ, respectively.

We combine the signal and counter-propagating control fields on a dichroic mirror. Both beams are focused down to a ~ 300 μm waist and temporally overlapped inside a 72 mm -long uncoated glass Cs-133 vapour cell heated to $\sim 91^\circ\text{C}$.

After the signal field leaves the memory it is sent into a Hanbury-Brown-Twiss detection setup, composed of a balanced beamsplitter and two fibre-coupled single-photon silicon APDs connected to a time-to-digital converter (same as the idler detector), allowing us to reconstruct the quantum photon number statistics of the stored/retrieved signal fields.

Photon statistics

By evaluating the $g^{(1,1)}$ cross-correlation function between the signal and idler (herald) pulses, we obtain

a measure for the strength of the correlations between them. $g^{(1,1)}$ is defined as $p_{\text{si}}/p_{\text{s}}p_{\text{i}}$, where p_{si} is the probability for a signal-idler coincidence click, and $p_{\text{s(i)}}$ is the signal (idler) click probability. Values of $g^{(1,1)} > 2$ signify non-classical correlations. To calculate $g^{(1,1)}$ from the measurements, we use

$$g^{(1,1)} = \frac{R_{\text{s,i}}}{R_{\text{s}}R_{\text{i}}}R_{\text{T}}, \quad (1)$$

where $R_{\text{s,i}}$ is the sum of $D_{\text{i}}\text{-}D_{\text{s1}}$ and $D_{\text{i}}\text{-}D_{\text{s2}}$ coincidences, R_{T} is the total number of trigger events during the whole measurement time, R_{s} is the sum of D_{s1} and D_{s2} clicks, and R_{i} is the number of D_{i} clicks.

The heralded auto-correlation is defined as $g_{\text{h}}^{(2)} = p_{(\text{s1},\text{s2}|\text{i})}/p_{(\text{s1}|\text{i})}p_{(\text{s2}|\text{i})}$, where $p_{(\text{s1},\text{s2}|\text{i})}$ is the conditional probability of detecting a coincidence between D_{s1} and D_{s2} given a click in D_{i} , and $p_{(\text{s1}|\text{i})}$ ($p_{(\text{s2}|\text{i})}$) is the probabilities to detect a click in D_{s1} (D_{s2}) given a click in D_{i} . Any $g_{\text{h}}^{(2)} < 1$ verifies non-classical photon-number statistics. We evaluate $g_{\text{h}}^{(2)}$ using

$$g_{\text{h}}^{(2)} = \frac{R_{\text{trip}}}{R_{\text{s1,i}}R_{\text{s2,i}}}R_{\text{i}}, \quad (2)$$

where R_{trip} is the number of triple coincidences between D_{i} , D_{s1} , and D_{s2} , R_{i} is the number of idler clicks, and $R_{\text{s1(2),i}}$ is the number of $D_{\text{i}}\text{-}D_{\text{s1}}$ ($D_{\text{i}}\text{-}D_{\text{s2}}$) coincidences. More details can be found in the Supplementary Material.

ACKNOWLEDGEMENTS

We would like to thank R. Chrapkiewicz and M. Parniak for useful discussions. We are grateful to H. Chrzanowski for proof-reading the manuscript.

This work was supported by the UK Engineering and Physical Sciences Research Council through Standard Grant No. EP/J000051/1, Programme Grant No. EP/K034480/1, and the EPSRC NQIT Quantum Technology Hub. We acknowledge support from the Air Force Office of Scientific Research: European Office of Aerospace Research and Development (AFOSR EOARD Grant No. FA8655-09-1-3020). J.N. acknowledges a Royal Society University Research Fellowship, and DJS acknowledges an EU Marie-Curie Fellowship No. PIFI-GA-2013-629229. P.M.L. acknowledges a European Union Horizon 2020 Research and Innovation Framework Programme Marie Curie individual fellowship, Grant Agreement No. 705278, and B.B. acknowledges funding from the European Unions Horizon 2020 research and innovation programme under grant

agreement No 665148. I.A.W. acknowledges an ERC Advanced Grant (MOQUACINO). S.E.T. and J.H.D.M. are supported by EPSRC via the Controlled Quantum Dynamics CDT under Grants EP/G037043/1 and EP/L016524/1. G.S.T. acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC). E.P. acknowledges an EU Marie-Curie Fellowship No. PIFI-GA-2013-627372. K.T.K. acknowledges a Santander Graduate Scholarship from Lady Margaret Hall, Oxford. O.L.-A. acknowledges support from a CONACYT scholarship.

AUTHOR CONTRIBUTIONS

K.T.K., A.F., E.P. and J.N. invented the protocol. K.T.K., G.S.T., O.L.-A., A.F. and J.N. developed the theory. K.T.K. designed the experiment. K.T.K., P.M.L., B.B., S.E.T., J.H.D.M. and D.J.S. built the setup. K.T.K. and P.M.L. conducted the experiment. B.B. wrote the data acquisition and analysis software. D.J.S., J.N. and I.A.W. supervised the work. All authors contributed to writing the manuscript.

-
- [1] H. J. Kimble. The quantum internet. *Nature*, 453(7198):1023–1030, 2008.
 - [2] Y. Li, P. C. Humphreys, G. J. Mendoza, and S. C. Benjamin. Resource costs for fault-tolerant linear optical quantum computing. *Physical Review X*, 5(4):1–15, 2015.
 - [3] C. Simon, M. Afzelius, J. Appel, A. Boyer de la Giroday, S. J. Dewhurst, N. Gisin, C. Y. Hu, F. Jelezko, S. Kröll, J. H. Müller, J. Nunn, E. S. Polzik, J. G. Rarity, H. de Riedmatten, W. Rosenfeld, A. J. Shields, N. Sköld, R. M. Stevenson, R. Thew, I. A. Walmsley, M. C. Weber, H. Weinfurter, J. Wrachtrup, and R. J. Young. Quantum memories. *The European Physical Journal D*, 58(1):1–22, April 2010.
 - [4] J. Nunn, N. K. Langford, W. S. Kolthammer, T. F. M. Champion, M. R. Sprague, P. S. Michelberger, X.-M. Jin, D. G. England, and I. A. Walmsley. Enhancing Multiphoton Rates with Quantum Memories. *Physical Review Letters*, 110(13):133601, mar 2013.
 - [5] T. Chanelière, D. N. Matsukevich, S. D. Jenkins, S. Y. Lan, T. A. B. Kennedy, and A. Kuzmich. Storage and retrieval of single photons transmitted between remote quantum memories. *Nature*, 438(7069):833, 2005.
 - [6] S. Zhou, S. Zhang, C. Liu, J. F. Chen, J. Wen, M. M. T. Loy, G. K. L. Wong, and S. Du. Optimal storage and retrieval of single-photon waveforms. *Optics express*, 20(22):24124–31, 2012.
 - [7] D. Maxwell, D. Szwert, D. Paredes-Barato, H. Busche, J. Pritchard, A. Gauguier, K. Weatherill, M. Jones, and C. Adams. Storage and Control of Optical Photons Using Rydberg Polaritons. *Physical Review Letters*, 110(10):103001, mar 2013.
 - [8] D. S. Ding, K. Wang, W. Zhang, S. Shi, M. X. Dong, Y. C. Yu, Z. Y. Zhou, B. S. Shi, and G. C. Guo. Entanglement between Lowly and Highly Lying Atomic Spin Waves. *Physical Review A*, 1:1–5, 2016.
 - [9] E. Distant, P. Farrera, A. Padrón-Brito, D. Paredes-Barato, G. Heinze, and H. de Riedmatten. Storing single photons emitted by a quantum memory on a highly excited Rydberg state. *Nature Communications*, 8:14072, 2017.
 - [10] D. G. England, K. A. G. Fisher, J.-P. W. MacLean, P. J. Bustard, R. Lausten, K. J. Resch, and B. J. Sussman. Storage and Retrieval of THz-Bandwidth Single Photons Using a Room-Temperature Diamond Quantum Memory. *Physical Review Letters*, 114(5):053602, 2015.
 - [11] B. Julsgaard, J. Sherson, J. I. Cirac, J. Fiurásek, and E. S. Polzik. Experimental demonstration of quantum memory for light. *Nature*, 432(7016):482–486, 2004.
 - [12] S. E. Thomas, J. H. D. Munns, K. T. Kaczmarek, C. Qiu, B. Brecht, A. Feizpour, P. M. Ledingham, I. A. Walmsley, J. Nunn, and D. J. Saunders. High Efficiency Raman Memory by Suppressing Radiation Trapping, arXiv:1610.03743 [quant-ph], 2016.
 - [13] F. Bussi eres, N. Sangouard, M. Afzelius, H. de Riedmatten, C. Simon, and W. Tittel. Prospective applications of optical quantum memories. *Journal of Modern Optics*, 60(18):1519–1537, 2013.
 - [14] M. D. Eisaman, A. Andr e, F. Massou, M. Fleischhauer, A. S. Zibrov, and M. D. Lukin. Electromagnetically induced transparency with tunable single-photon pulses. *Nature*, 438(7069):837–41, 2005.
 - [15] M. Hosseini, G. Campbell, B. M. Sparkes, P. K. Lam, and B. C. Buchler. Unconditional room-temperature quantum memory. *Nature Physics*, 7(10):794–798, 2011.
 - [16] P. Jobez, C. Laplane, N. Timoney, N. Gisin, A. Ferrier, P. Goldner, and M. Afzelius. Coherent Spin Control at the Quantum Level in an Ensemble-Based Optical Memory. *Physical Review Letters*, 114(23):1–5, 2015.

- [17] M. P. Hedges, J. J. Longdell, Y. Li, and M. J. Sellars. Efficient quantum memory for light. *Nature*, 465(7301):1052–6, June 2010.
- [18] A. Seri, A. Lenhard, D. Rieländer, M. Gündoğan, P. M. Ledingham, M. Mazzera, and H. De Riedmatten. Quantum correlations between single telecom photons and a multimode on-demand solid state quantum memory, arXiv:1701.09004v1 [quant-ph], 2017.
- [19] K. F. Reim, J. Nunn, V. O. Lorenz, and B. J. Sussman. Towards high-speed optical quantum memories. *Nature Photonics*, (March):8–11, 2010.
- [20] R. Chrapkiewicz, M. Dąbrowski, and W. Wasilewski. High-Capacity Angularly Multiplexed Holographic Memory Operating at the Single-Photon Level. *Physical Review Letters*, 118(6):063603, 2017.
- [21] K. F. Reim, J. Nunn, X.-M. Jin, P. S. Michelberger, T. F. M. Champion, D. G. England, K. C. Lee, W. S. Kolthammer, N. K. Langford, and I. A. Walmsley. Multipulse Addressing of a Raman Quantum Memory: Configurable Beam Splitting and Efficient Readout. *Physical Review Letters*, 108(26):263602, jun 2012.
- [22] G. T. Campbell, O. Pinel, M. Hosseini, T. C. Ralph, B. C. Buchler, and P. K. Lam. Configurable unitary transformations and linear logic gates using quantum memories. *Physical Review Letters*, 113(6):1–5, 2014.
- [23] K. A. G. Fisher, D. G. England, J.-P. W. MacLean, P. J. Bustard, K. J. Resch, and B. J. Sussman. Frequency and bandwidth conversion of single photons in a room-temperature diamond quantum memory. *Nature communications*, 7:11200, 2016.
- [24] P. S. Michelberger, T. F. M. Champion, M. R. Sprague, K. T. Kaczmarek, M. Barbieri, X. M. Jin, D. G. England, W. S. Kolthammer, D. J. Saunders, J. Nunn, and I. A. Walmsley. Interfacing GHz-bandwidth heralded single photons with a room-temperature Raman quantum memory. *New Journal of Physics*, 17(4):43006, 2015.
- [25] J. Nunn, I. A. Walmsley, M. Raymer, K. Surmacz, F. Waldermann, Z. Wang, and D. Jaksch. Mapping broadband single-photon wave packets into an atomic memory. *Physical Review A*, 75(1):011401, jan 2007.
- [26] In preparation.
- [27] M. Gündoğan, P. M. Ledingham, K. Kutluer, M. Mazzera, and H. De Riedmatten. Solid State Spin-Wave Quantum Memory for Time-Bin Qubits. *Physical Review Letters*, 114(23):1–5, 2015.
- [28] I. Novikova, A. V. Gorshkov, D. F. Phillips, A. S. Sørensen, M. D. Lukin, and R. L. Walsworth. Optimal control of light pulse storage and retrieval. *Physical Review Letters*, 98(24):1–4, 2007.
- [29] B. Huber, T. Baluktsian, M. Schlagmüller, A. Kölle, H. Kübler, R. Löw, and T. Pfau. GHz Rabi Flopping to Rydberg States in Hot Atomic Vapor Cells. *Physical Review Letters*, 107(24):243001, December 2011.
- [30] N. Ph. Georgiades, E. S. Polzik, and H. J. Kimble. Two-photon spectroscopy of the $6S_{1/2}$ - $6D_{5/2}$ transition of trapped atomic cesium. *Optics Letters*, 19(18):1474–1476, 1994.

SUPPLEMENTARY MATERIAL

S.1. Memory lifetime

The two main limiting factors to the ORCA memory lifetime are spontaneous emission from the doubly-excited storage state and motional-induced dephasing due to the wavevector mismatch of the signal and control fields. In the case of storage into the $6D_{5/2}$ manifold in Cs, as demonstrated in this paper, the expected lifetime of the memory due to spontaneous emission would be ~ 51 nanoseconds [30].

Motional-induced dephasing arises due to the Doppler effect and atomic motion in a warm ensemble. This is because the stored excitation is spread over atoms belonging to different velocity classes in the ensemble. Due to the Doppler effect, each velocity class experiences light with a different frequency. As a consequence, the phase of the stored coherence in different velocity classes will evolve at different rates. Therefore the collective excitation will dephase at a rate proportional to the range of atomic velocities. The motional-induced dephasing lifetime is:

$$\tau_D = \sqrt{\frac{3}{2}} \frac{1}{k_r v_D}, \quad (\text{S.1})$$

where v_D is the rms velocity of the atoms and $k_r = \frac{2\pi}{\lambda_s} - \frac{2\pi}{\lambda_c}$ is the wavevector of the stored coherence (λ_s and λ_c are the signal and control wavelengths, respectively). In our case, $v_D \approx 261$ m/s (inferred from the ensemble temperature) and $\lambda_s = 852.35$ nm, $\lambda_c = 917.47$ nm, yielding $\tau_D \approx 8.96$ nanoseconds. Since 8.96 ns $\ll 51$ ns, the memory lifetime should be currently limited by motional-induced dephasing, i.e. an inhomogenous process. Therefore we can assume a Gaussian-shaped decay of the memory efficiency with time.

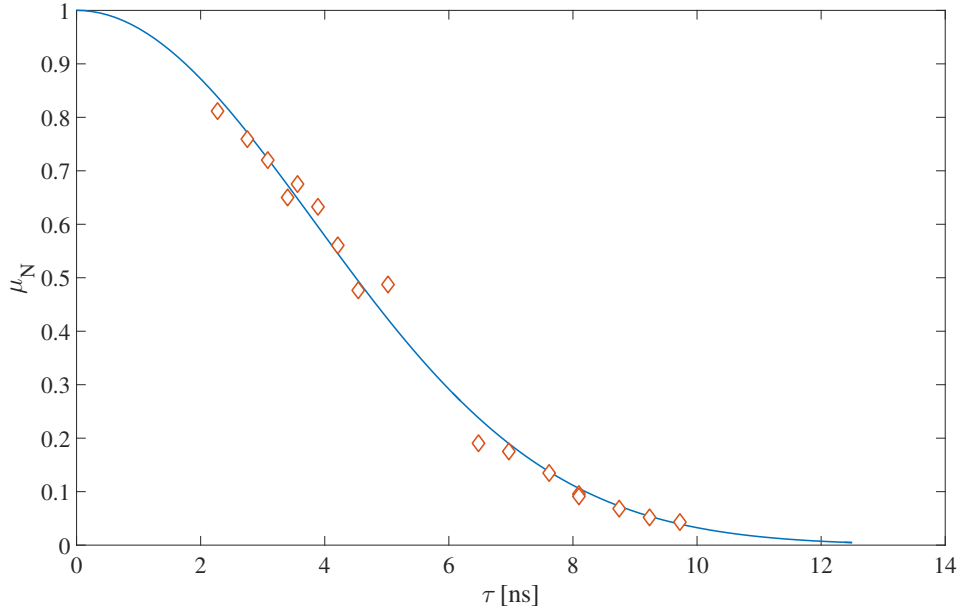


FIG. S.1. Memory lifetime curve. The measured normalised (to $\tau = 0$) memory efficiency η_N (orange diamonds) versus storage time τ along with a fit of the memory lifetime curve (blue line) yielding a lifetime of 5.4 ± 0.1 ns (experimental errors are smaller than the markers).

We extract the actual memory lifetime by measuring the memory efficiency for different storage times τ using a weak coherent state signal. On Fig. S.1 we show the measured normalised (to $\tau = 0$) memory efficiency along with a Gaussian fit to the data, obtaining a $(1/e)$ lifetime of 5.4 ± 0.1 ns, close to our estimate of 8.96 ns. We attribute the discrepancy to storage in different hyperfine states in the $6D_{5/2}$ manifold. These are spread over ~ 100 MHz, with different phase acquisition rates, leading to an oscillation in memory efficiency on a timescale of ~ 11 ns.

While seemingly short, this lifetime can be further extended by e.g. using a different atomic species with a smaller signal/control wavelength mismatch. For example, applying Eq. (S.1) to the $5S_{1/2} \rightarrow 5P_{3/2} \rightarrow 5D_{5/2}$ cascade in rubidium (signal at 780 nm, control at 776 nm), we predict a motional-induced dephasing lifetime of $\tau_D \approx 90$ ns.

S.2. Experimental setup

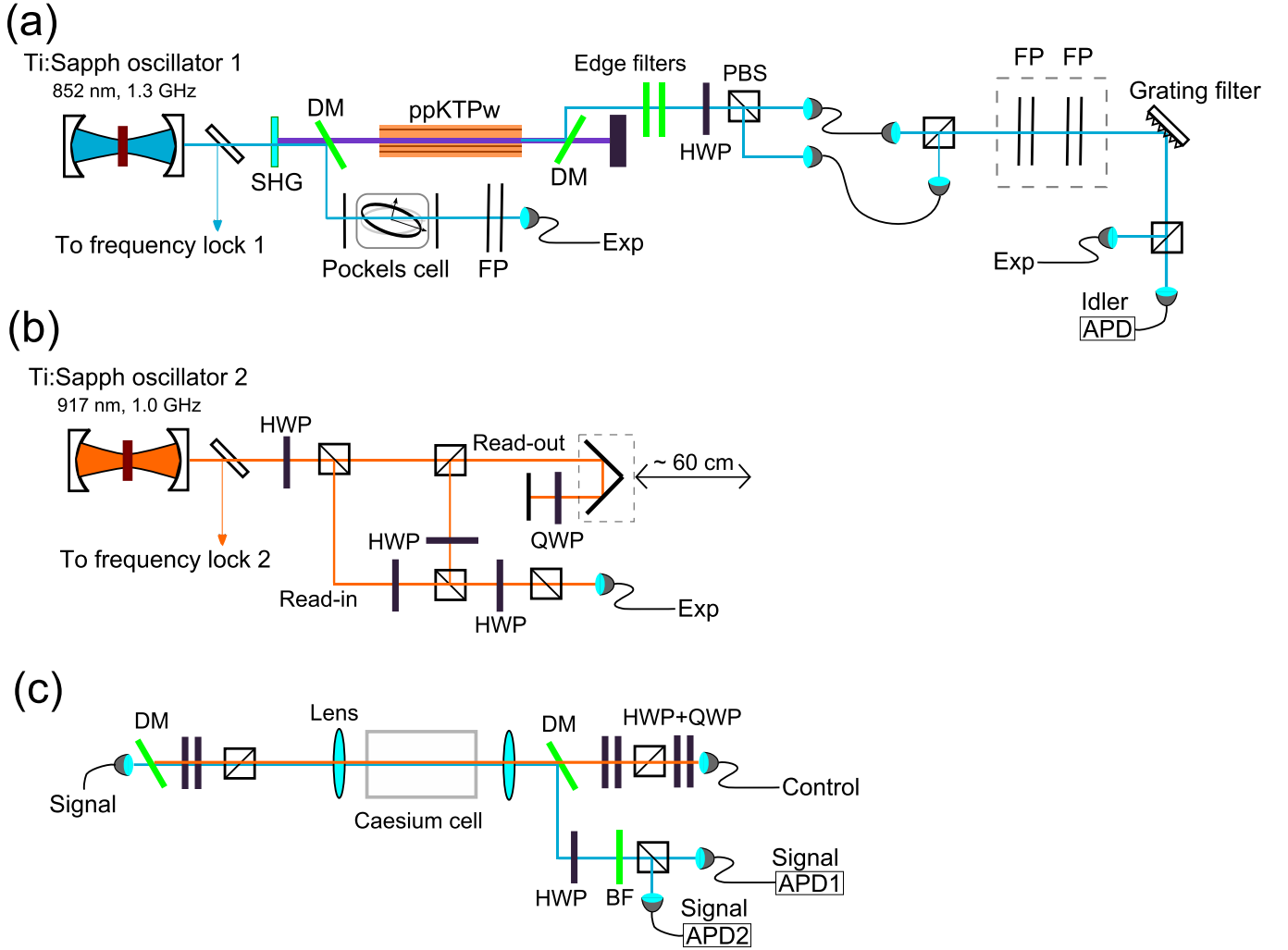


FIG. S.2. Schematic of the experimental setup for ORCA. (a) Signal generation stage. (b) Control generation stage. (c) Memory and detection stage. SHG - periodically poled potassium titanyl (ppKTP) bulk crystal; ppKTPw - ppKTP waveguide; DM - dichroic mirror; FP - Fabry-Perot etalon; PBS - polarizing beamsplitter; HWP - half-wave plate; QWP - quarter-wave plate; APD - avalanche photodiode detector; BF - bandpass filter.

Figure S.2 shows a schematic of the ORCA experimental setup. For the memory lifetime measurement we produce a weak coherent state signal (average photon number of 2) by picking pulses using a fast Pockels cell (extinction 20,000:1) at a 1 MHz rate from a 80 MHz train of pulses generated by a ~ 330 ps actively mode-locked titanium sapphire (Ti:Sapph) laser operated at 852 nm and filtered by a Fabry-Perot (FP) etalon down to ~ 1 GHz bandwidth. Using a scanning FP etalon connected to a PC running LabVIEW, we reference-lock the Ti:Sapph's center frequency (via the voltage on a Gires-Tournois-Interferometer inside the laser cavity) to a continuous wave (CW) laser locked to the Cs D2 line via saturated absorption spectroscopy.

We generate the control field from a second ~ 440 ps actively mode-locked Ti:Sapph laser operated at 917 nm, with its center frequency locked using a wavelength meter and its repetition rate locked to the signal Ti:Sapph using a commercial lock-to-clock (L2C) system. We use an unbalanced free-space Mach-Zehnder interferometer to split the 80 MHz pulse train into two, with a variable delay $< \sim 4$ ns between them, in order to investigate storage times < 12.5 ns. For each storage time $\sim 6 < \tau < 12.5$ ns we use the L2C electronics to change the timing between signal and control pulse trains such that read-in and read-out are switched. We also use the L2C to temporally overlap the signal and control pulses in the memory cell.

We combine the signal and control fields on a dichroic mirror, which - followed by a 10 nm bandpass filter centred at 850 nm - reduces control field leakage to the detectors from back-reflections by a factor of $\sim 10^9$. We focus signal

and control beams down to a $\sim 300 \mu\text{m}$ waist inside a 72 mm -long uncoated caesium borosilicate reference cell heated using a custom-made oven. We estimate the cell temperature to be $\sim 91^\circ\text{C}$ by frequency scanning a weak CW probe laser over the Cs D2 line and fitting a Voigt profile to the measured atomic absorption line.

After the signal field passes through the memory and the filters, we send it into a Hanbury-Brown-Twiss setup, composed of a half-waveplate, polarising beamsplitter and two fibre-coupled single-photon avalanche photodiodes. The two signal and the idler avalanche photodiodes were connected to a time-to-digital converter. For the weak coherent state data, we add the counts on the two signal detectors to estimate the total magnitude of the transmitted/retrieved signal.

S.3. Photon Source

The generation of heralded single photons is achieved using type-II parametric down-conversion in a ppKTP waveguide. The waveguide, operated in a single-pass configuration and of length 20 mm, is pumped with pulses of approximately 260 ps duration at a wavelength of 426 nm. This pump light is derived by doubling the above mentioned 852 nm Ti:Sapph laser via second harmonic generation in a separate 2 mm long ppKTP crystal. With an incident average power near 700 mW at 852 nm at the crystal we arrive with 4 mW average power at 426 nm before the PDC waveguide. This light is then coupled to the waveguide with a total transmission of $< 10\%$ including the loss at the in- and out-coupling lenses. We note that the waveguide is not single-mode for the pump wavelength and that the coupling is optimised to primarily excite the fundamental spatial mode, resulting in a low overall transmission. The generated frequency-degenerate but polarisation-orthogonal signal and idler modes have a bandwidth on the order of 1 THz.

We characterise beam propagation transmission using an “alignment” mode, which is coupled to the fundamental mode of the waveguide, and thereby comparable to the signal and idler modes allowing for “classical” measurements to be made. These modes are then subject to frequency filtering. First, we apply coarse filtering using edge filters. Then, the modes are spatially separated via a PBS to then be coupled to their own single-mode fibre (SMF) with an efficiency of $(64 \pm 1)\%$ for the signal mode and $(53 \pm 2)\%$ for the idler mode. The modes are then out-coupled and recombined on a PBS forming a common spatial mode to then pass two etalons, one of $\text{FSR} = 18.4 \text{ GHz}$ and one of 103 GHz , which gives an effective FSR of 1 THz (lowest common multiple). This is followed by a holographic volume Bragg grating with a width $\sim 100 \text{ GHz}$. Using a narrowband probe ($\sim \text{MHz}$) the measured width this filtering has is 1 GHz for both modes. Finally the modes are separated spatially again via a PBS, the idler coupled to an APD (efficiency $\eta \approx 50\%$, dark counts = $163 \pm 1 \text{ Hz}$) via a multimode fibre (total transmission from after waveguide to in front of idler detector is $\eta_{i,\text{filt}} = (9.7 \pm 0.1)\%$), while the signal is coupled to a SMF to be out-coupled and steered to the memory (total transmission from after waveguide to after this SMF is $\eta_{s,\text{filt}} = (12.8 \pm 0.3)\%$).

The filtered signal photon is now steered toward the memory. First there is an edge filter which is used to prevent the 917 nm control from backward-coupling toward the source which presents additional loss to the signal mode. Further, the caesium cell used is uncoated, adding more loss. After passing the cell, the signal mode is then separated from the control mode via a dichroic mirror and finally passes a bandpass filter about 852 nm before entering a Hanbury-Brown-Twiss set-up. The mode is spatially separated into two and coupled to two APDs ($\eta \approx 50\%$ dark counts = $296 \pm 2 \text{ Hz}$ and $\eta \approx 50\%$ dark counts = $356 \pm 2 \text{ Hz}$) via SMF. The total transmission from the source to in front of these detectors (averaging over the two SMF couplings) is $\eta_{s,\text{total}} = 3.7 \pm 0.1\%$. That is to say, the photon undergoes an additional $\eta_{s,\text{add}} = 30\%$ transmission from after the initial filtering stage.

For all results presented in this paper we operated with an average pump power of 4 mW in front of the waveguide in-coupling lens. Typically, we measure an idler (signal) count rate of around 30 kHz (10 kHz) for the case of no control pulses. The typical Klyshko efficiency η_k measured is 0.7%. This allows to calculate a heralding efficiency of $\eta_{\text{herald}} = \eta_k / \eta_{\text{det}} / \eta_{s,\text{add}} = 4.7\%$, which is well above the μ_1 of the memory. Finally, the heralding efficiency just after the waveguide is $\eta_{s,\text{waveguide}} = \eta_k / \eta_{\text{det}} / \eta_{s,\text{total}} = 48\%$. The missing factor of 2 we attribute to not measuring explicitly the loss inside the waveguide, the out-coupling loss from waveguide to free-space and the potential frequency mismatch of the etalon centroids between signal and idler.

S.4. Data acquisition and post-processing

During the measurements, the settings of three mechanical shutters which selectively blocked the read-in, read-out, and signal beams defined four different configurations (see Fig. S.3): memory measurements with all three shutters open (MEM); read-in measurements with signal and read-in shutters open, and read-out shutter closed (RI); signal measurements with signal shutter open and both read-in and read-out closed (SIG); and noise measurements with read-in and read-out shutters open and signal shutter closed (CTRL). A single measurement consisted of recording

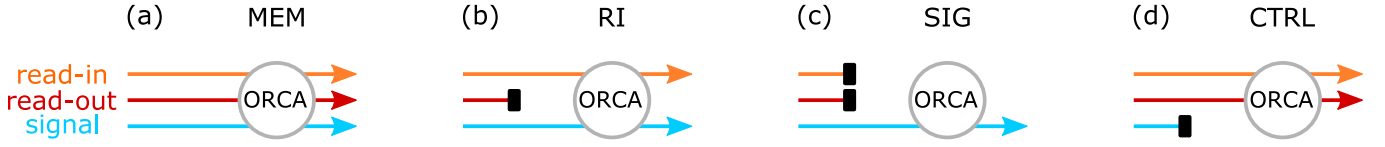


FIG. S.3. Schematic of the different measurement settings. Black rectangles signify a closed shutter. For more information see the text.

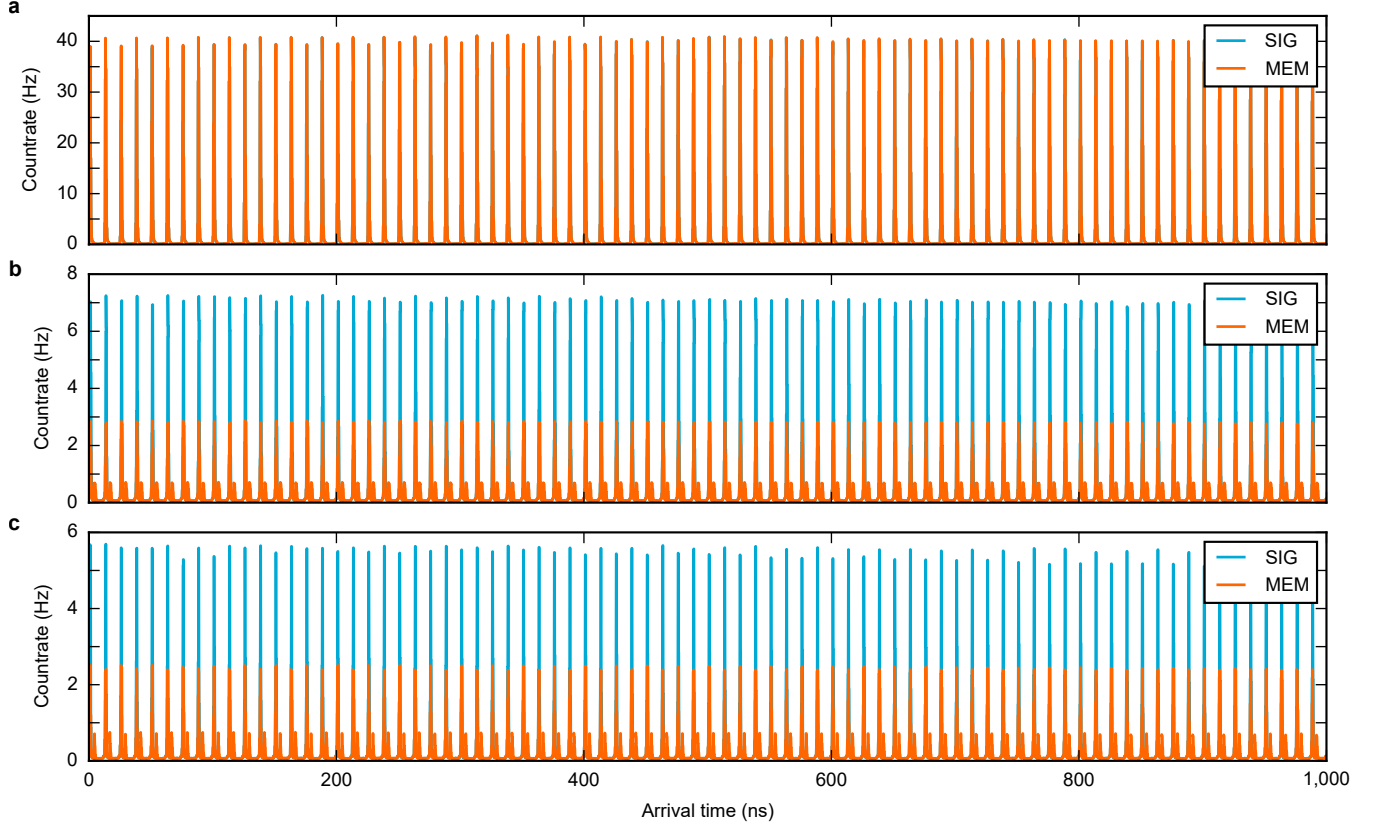


FIG. S.4. **Accumulated normalised arrival time histograms.** **a** Histogram of firing times of detector D_i with respect to a 1 MHz trigger derived from the Ti:Sapph for SIG (blue trace) and MEM (orange trace) configuration. **b** The same histograms for detector D_{s1} . The reduced height of the main peaks in the MEM configuration signifies read-in that is, storage of photons. The additional peaks 3.5 ns after the main peaks originate from retrieved light that has been read out from the memory. **c** The same as in **b**, just for detector D_{s2} .

the number of detector counts registered in a period of 180 s in the MEM configuration, followed by recording the total counts over 10 s in the RI, SIG, and CTRL configurations. After completion of all four configurations, the corresponding data was written to disk and the measurement repeated. This mode of operation was chosen in order to mitigate the effect of slow drifts in the setup that arose from changes in the laboratory temperature and humidity.

For each configuration in each measurement, we recorded arrival time histograms for the three detectors (D_i , D_{s1} , D_{s2}). These are histograms of firing times of the single-photon detectors with respect to a 1 MHz trigger signal derived from the Ti:Sapph recorded with the time-to-digital (TDC) converter. We chose a time-bin width of 200 ps as a compromise between temporal resolution of the TDC and total number of time bins in the histogram. For data visualisation, we added all arrival time histograms and normalised them to both the number of measurements (521) and the respective measurement duration (180 s for MEM, 10 s else), obtaining a count rate per time bin in units of Hertz. This allows for an easy comparison of different measurement configurations with potentially different total measurement times. Fig. S.4 shows the resulting accumulated arrival time histograms for D_i , D_{s1} , and D_{s2} (a to c) and the SIG (blue trace) and MEM (orange trace) configurations. Note that these are unprocessed raw data, which demonstrate the high signal-to-noise ratio of the measurements, signified by the good visibility of the peaks.

Fig. S.5 shows a section of the histograms from Fig. S.4. To reduce the impact of spurious noise counts (primarily

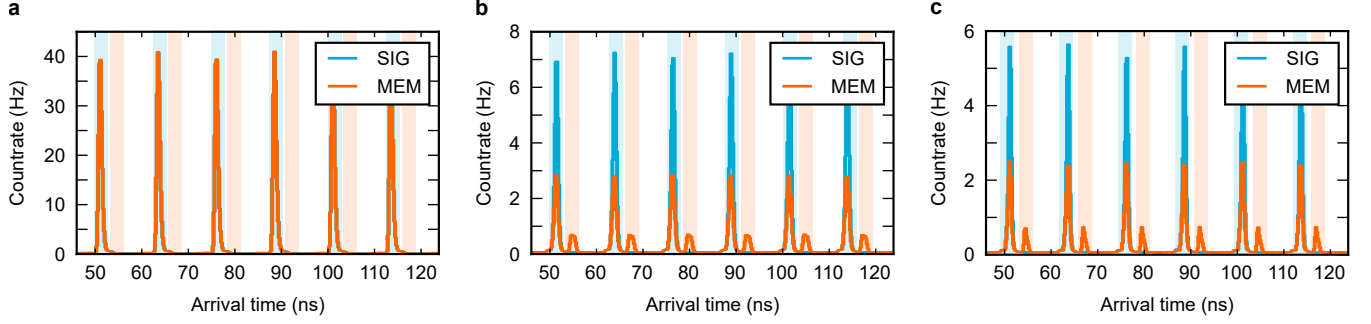


FIG. S.5. **Sections of the arrival time histograms with indicated time gates.** **a** A section of the arrival time histogram for D_i from Fig. S.4a, with indicated time gates. The blue-shaded regions correspond to read-in peaks, whereas the orange-shaded regions symbolize read-out peaks. For data evaluation, only events that lay within the time gates were retained, in order to suppress the contributions of spurious background counts. Again, we show histograms for the SIG (blue trace) and MEM (orange trace) configurations. **b** A section of the arrival time histograms for detector D_{s1} . **c** The same for detector D_{s2} .

from detector dark counts), we applied time gates to the recorded arrival time histograms and only retained events that lay within the time gates. The time gates for the read-in pulses (blue regions) are centred around the maxima of the individual read-in peaks and have a width of 2.5 ns, chosen such that the peaks were completely inside the gating region. Similar time gates were chosen for the retrieved light, where the centre of these read-out gates (orange regions) was offset from the corresponding read-in time gates by 3.5 ns, which was the storage time chosen for the experiment. By integrating the detection events over only the gate regions, we calculated the read-in, read-out, and total memory efficiency as stated in the main text.

In addition to the arrival time histograms, we also recorded coincidence histograms for signal-idler two-fold coincidences (D_i - D_{s1} , D_i - D_{s2}) and the three-fold coincidences (D_i - D_{s1} - D_{s2}). These are start-stop histograms, where the detection of an idler photon starts the measurement and the detection of a signal photon (the detection of a D_{s1} - D_{s2} coincidence) serves as the stop signal for the two-fold (three-fold) coincidence measurement. In this case, the time-bin width of the TDC was chosen to be 100 ps to ensure that the temporal resolution of the measurement was not limited by the TDC. Again, data was recorded for each configuration in each measurement, and the resulting histograms were accumulated and normalised with respect to the number of measurements and the respective measurement duration, yielding coincidence rates per time bin in units of Hertz for data visualisation.

The resulting coincidence traces are plotted in Fig. ??, where we show D_i - D_{s1} -, D_i - D_{s2} -, and D_i - D_{s1} - D_{s2} -coincidences (top to bottom) for both SIG (blue traces) and MEM (orange traces) configurations. Note that the unconventional shape of the traces originates from the logarithmic scaling of the y axes.

The SIG traces show a dominant peak at a time difference of 0 ns, with subsequent smaller peaks at integer multiples of the laser repetition time of 12.5 ns. The ratio of the area under the main peak and the area of a side peak, the so-called coincidences-to-accidentals ratio (CAR), evaluates to $\text{CAR} \approx 130$. This number is also a measure for the amount of higher-order photon-number components in our PDC. The main peak originates from correlated signal and idler photons that have been generated in the same PDC process. Therefore, if the probability to generate a photon pair is given by p , we find that the area under the main peak is proportional to p . In contrast, the subsequent peaks originate from two uncorrelated PDC events, where photons have been generated by subsequent laser pulses, and consequently we find that the area of a side peak is proportional to p^2 . From this we estimate $p \approx 7.7 \cdot 10^{-3} \approx 1/130$, which puts us into the low-gain regime of PDC. In this case, the probability to generate two photon pairs in one single PDC process is approximately given by p^2 , and we conclude that we primarily generate single photon pairs during the PDC.

This can further be quantified by calculating the $g^{(1,1)}$ signal-idler cross-correlation function for the SIG configuration. The results of this are summarised in the first row of Tab. ??, where we find $g^{(1,1)} = 131(1)$ for a time difference of 0 ns and $g^{(1,1)} \approx 1$ for integer multiples of 12.5 ns. A value of $g^{(1,1)} > 2$ signifies non-classical correlations between signal and idler photons, and we find that indeed these are only present for photons that originate from the same PDC process. We also note that the values at the read-out times (3.5 ns offset from the 12.5 ns time slots) are meaningless, since there is no actual signal at the detectors.

Turning our attention to the MEM configuration (orange traces in Fig. ??), we again find a dominating peak at a time difference of 0 ns with side peaks at integer multiples of 12.5 ns. In addition, we see peaks that are offset from the major peaks by 3.5 ns. These originate from coincidence events between idler photons and signal photons that have been stored in and retrieved from the memory. We also note that the side peak at 12.5 ns is higher than the corresponding peak for the SIG configuration. The reason for this lies in the non-unity read-out efficiency of our

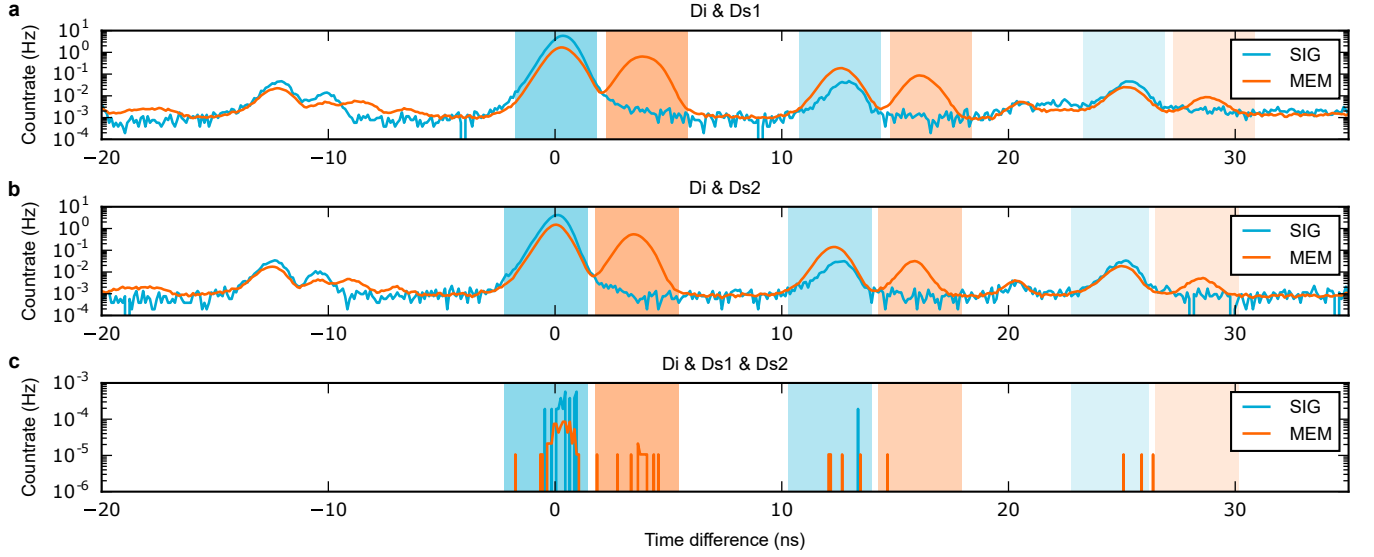


FIG. S.6. **Correlation histograms.** **a** Accumulated normalised correlation histogram for two-fold coincidences between detectors D_i & D_{s1} , shown for the SIG (blue trace) and MEM (orange trace) configurations. Note the logarithmic scaling of the y axis. We analyse the $g^{(1,1)}(0)$ cross-correlation for the initial time at 0 ns (read-in; blue-shaded) and 3.5 ns (read-out; orange-shaded). Successive read-in (read-out) time bins are visualised by shaded regions with decreasing saturation. **b** The same as in **a**, now however for two-fold coincidences between detectors D_i & D_{s2} . **c** Correlation histogram for three-fold coincidences between detectors D_i & D_{s1} & D_{s2} .

	$g^{(1,1)}(0)$					
	$\tau = 0$ ns	$\tau = 3.5$ ns	$\tau = 12.5$ ns	$\tau = 16$ ns	$\tau = 25$ ns	$\tau = 28.5$ ns
SIG	131(1)	0.203(1)	1.13(2)	0.0773(1)	1.23(2)	0.137(1)
MEM	80.5(1)	120(1)	9.74(2)	11.3(1)	1.78(2)	1.19(1)

TABLE S.1. **Cross-correlation for successive read-outs** Calculating $g^{(1,1)}(0)$ for higher-order read-outs at integer multiples of 12.5 ns (plus 3.5 ns for read-out pulses) yields the preservation of non-classical correlations by the memory up to around three times the memory lifetime.

memory. A stored photon is not necessarily read out after 3.5 ns, but can remain stored in the memory. Then, it can be read out by the next laser pulse arriving at 12.5 ns, and so on. To quantify this effect, we again evaluate the $g^{(1,1)}$ cross-correlation function. The results are summarised in the second row of Tab. ???. In this case, we find nonclassical values for $g^{(1,1)}$ up to a time of 16 ns, which corresponds to around three times the lifetime of our memory. These results highlight the noise-free operation of ORCA: non-classical photon correlations are retained even after the memory efficiency has decayed to around 5% of its initial maximum value ($1/e^3$).